



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

This hands-on guide explains how a Business Analyst without deep statistical know-how can identify customer contacts that are most likely to respond positively to a Marketing campaign.

SAP Predictive Analytics, Automated Mode is creating the predictive model that scores customers on their interest for a certain product. This analysis is based on the outcome of a previous Marketing campaign, which is stored locally in a flat file.

This guide is giving a high-level introductory overview and only shows a small fraction of the available functionality.

Since the data used in this guide is publicly available, the reader can follow the steps hands-on and carry out the same analysis.

October 2015

**Andreas Forster**

Predictive Presales Expert

SAP Switzerland

[andreas.forster@sap.com](mailto:andreas.forster@sap.com)

## TABLE OF CONTENTS

|  |    |
|--|----|
| INTRODUCTION .....                         | 3  |
| BUSINESS CASE .....                        | 3  |
| PRE-REQUISITES .....                       | 3  |
| DATA .....                                 | 4  |
| IMPLEMENTATION .....                       | 7  |
| Create Predictive Model .....              | 7  |
| Understand Predictive Model .....          | 15 |
| Apply Predictive Model .....               | 25 |
| SUMMARY .....                              | 29 |
| OPTIONAL MODIFICATIONS .....               | 30 |
| Model Performance on new Data .....        | 30 |
| Describe the customer in more detail ..... | 34 |
| Output customer ID with probability .....  | 37 |

### INTRODUCTION

SAP Predictive Analytics, Automated Mode enables data analysts to easily carry out high-quality, robust predictions with little effort. Tasks that typically require deep statistical expertise and high effort in conventional data mining environments are handled automatically. For instance SAP Predictive Analytics takes care of:

- Handling of missing values and outliers
- Binning of continuous variables
- Grouping of categories.
- Multicollinearity
- Column selection
- Model Selection

The user interface is designed for high usability. No expert scripting knowledge is required. Thanks to a high performance engine, SAP Predictive Analytics can handle very large data volumes. Models can be created on thousands of columns. No prior column selection has to be carried out. Any column that might possibly have an impact on the score can be included.

Once the model has been created, the user can look into the model to understand its findings and workings.

If you would like to understand how the best model is automatically found, you can see this [article](#)<sup>1</sup> for a high-level overview.

### BUSINESS CASE

The Marketing department of a bank wants to analyse the success of their last Marketing campaign to increase the response rate of the next campaign. In order to make the most of our Marketing budget we want to be able to address only those customers that are most likely to respond positively.

We create a model based on the historical data of our last campaign that indicates whether the customer did or did not purchase a product from a previous Marketing campaign. This model can then be applied on customers that have not yet been contacted. A probability score is calculated for each customer, which will indicate the most relevant customers for the campaign.

### PRE-REQUISITES

No previous knowledge about creating predictive models is required. Since we are connecting to a flat file, no database knowledge is required either.

You need to have SAP Predictive Analytics installed. At the time of writing, you can download an evaluation copy on [www.sap.com/trypredictive](http://www.sap.com/trypredictive).

---

<sup>1</sup> How does Automated Analytics do it? The magic behind creating predictive models automatically  
<http://scn.sap.com/docs/DOC-65046>

### DATA

We are using a dataset from a Portuguese bank that was shared publically as part of the following publication:

*[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014*

The bank carried out a marketing campaign by telephone to promote a term deposit. The dataset contains information about 45211 customers that were contacted, out of which 5289 did purchase the product.

The authors have kindly agreed that their data can be used for demonstration purposes by SAP. You can download the file bank-full-withid.csv on GitHub.<sup>2</sup>

If you would like to download the original “Bank Marketing Data Set”, please go to

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

Click into the “Data Folder” and download bank.zip, in which you will find the file bank-full.csv. This tutorial uses the same data, just with the addition of an identification column “row\_id”. This column will be used in the tutorial to identify the buying propensity of individual customers.

---

<sup>2</sup> <https://github.com/AndreasForster/Predictive/blob/master/AutomatedClassificationTutorial/bank-full-withid.csv>

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

The dataset is described as follows by the authors:

### Input Variables

| Number | Name / Description                                  | Type / Content  |
|--------|---|---|
| 1      | age   | Numeric   |
| 2      | job   | Categorical<br>'admin.', 'unknown', 'unemployed', 'management', 'housemaid', 'entrepreneur', 'student', 'blue-collar', 'self-employed', 'retired', 'technician', 'services' |
| 3      | marital   | Categorical<br>'married', 'divorced', 'single'; note: 'divorced' means divorced or widowed  |
| 4      | education   | Categorical<br>'unknown', 'secondary', 'primary', 'tertiary'  |
| 5      | default:<br>has credit in default?                  | Binary<br>'yes', 'no'   |
| 6      | balance:<br>average yearly balance, in euros        | Numeric   |
| 7      | housing:<br>has housing loan? (binary: 'yes', 'no') | Binary<br>'yes', 'no'   |
| 8      | loan:<br>has personal loan?                         | Binary<br>'yes', 'no'   |
| 9      | contact:<br>contact communication type              | Categorical<br>'unknown', 'telephone', 'cellular'   |
| 10     | day:<br>last contact day of the month               | Numeric   |
| 11     | month:<br>last contact month of year                | Categorical<br>'jan', 'feb', 'mar', ..., 'nov', 'dec'   |
| 12     | duration:<br>last contact duration, in seconds      | Numeric   |

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

### Other Attributes

| Number | Name / Description  | Type  |
|--------|---|---|
| 13     | campaign:<br>number of contacts<br>performed during this<br>campaign and for this client                      | Numeric   |
| 14     | pdays:<br>number of days that<br>passed by after the client<br>was last contacted from a<br>previous campaign | Numeric,<br>1 means client was not previously contacted |
| 15     | previous:<br>number of contacts<br>performed before this<br>campaign and for this client                      | Numeric   |
| 16     | poutcome:<br>outcome of the previous<br>marketing campaign  | Categorical:<br>'unknown','other','failure','success'   |

### Output Variable (Desired Target)

| Number | Name / Description                                 | Type                  |
|--------|--|-----------------------|
| 17     | y:<br>has the client subscribed a<br>term deposit? | Binary:<br>'yes','no' |

### IMPLEMENTATION

#### Create Predictive Model

Start SAP Predictive Analytics. This document uses version 2.2.

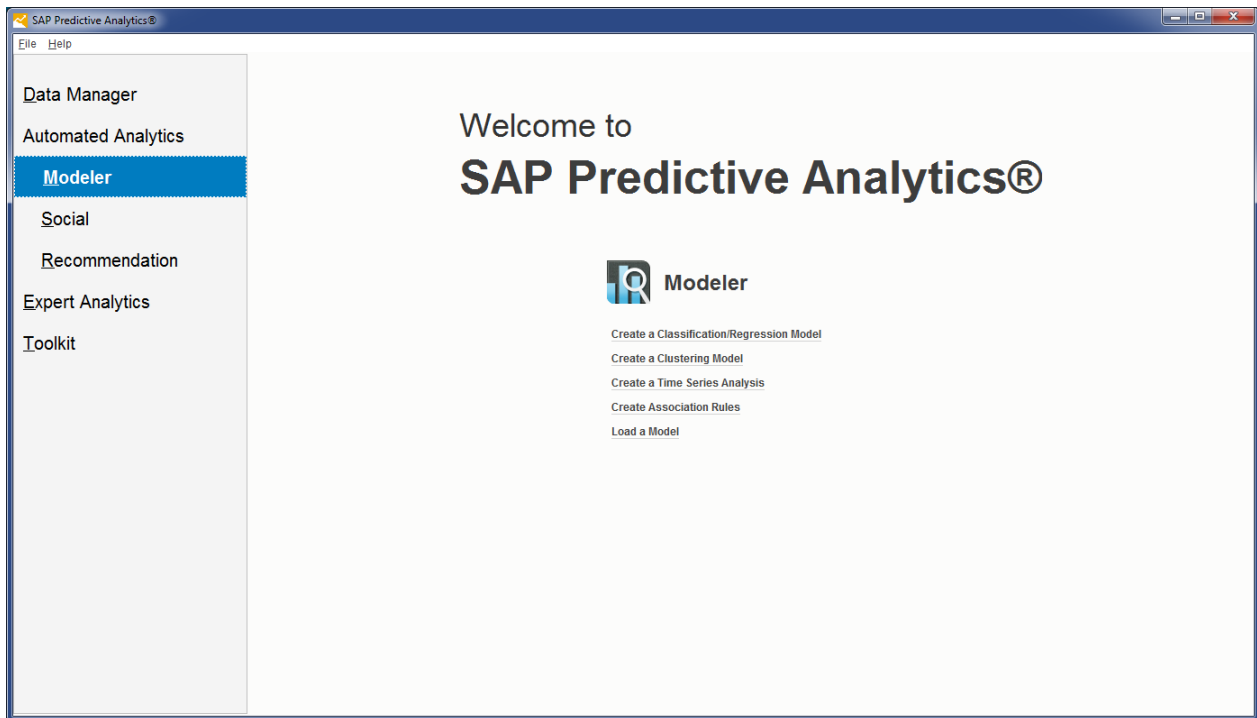


On the left hand-side you see the different categories of functionalities.

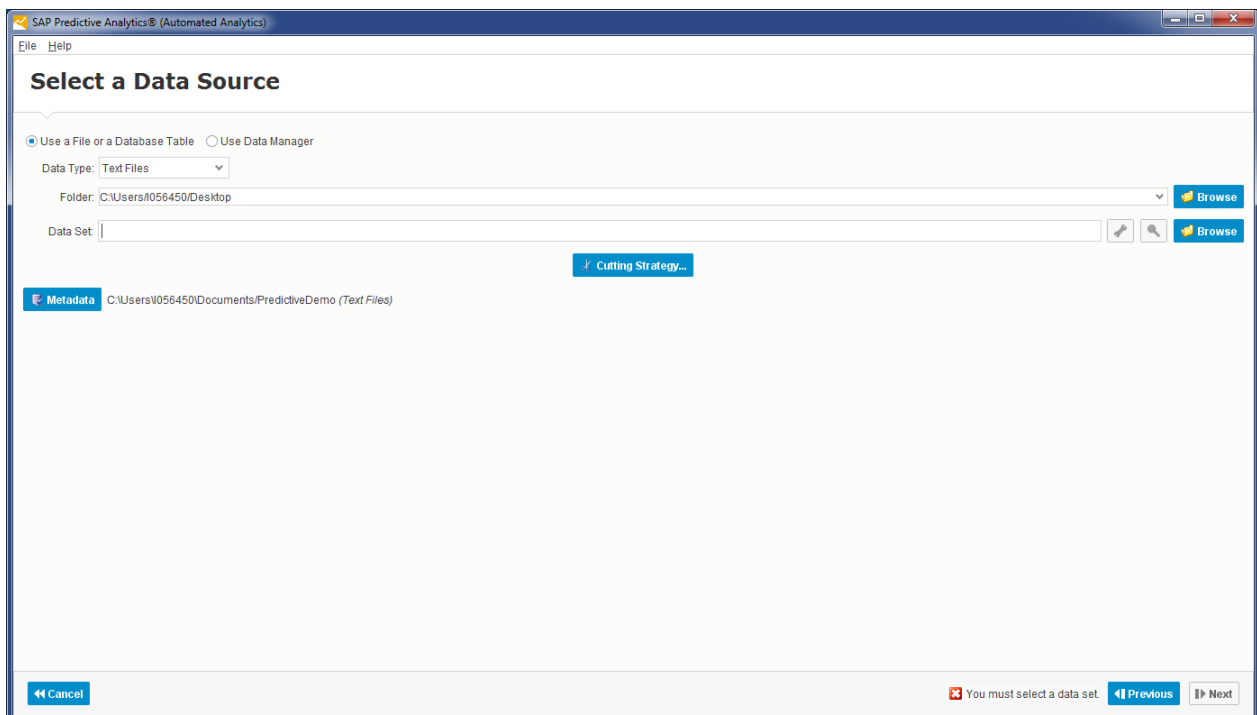
| Option              | Funcitonality   |
|---------------------|---|
| Data Manager        | Enriches the data, ie through the semantic creation of new variables, table joins, time references, filters and pivoting. |
| Automated Analytics | Automatically creates analytical models, ie to classify, score, cluster or to forecast.                                   |
| Expert Analytics    | Graphical workbench to build specific analytical workflows using individual algorithms/components.                        |
| Toolkit             | Complementary functionality such as previewing or transferring data.  |

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

The “Data Manager” functionality is optional. We start creating predictive models directly on the data file. Just click the “Modeler” option in the “Automated Analytics” category.



The identification of whether a person is interested in a product or not can be done with a classification. We classify the people in the two categories “yes” and “no”. Therefore click on the option “Create a Classification/Regression Model” in the center of the screen.

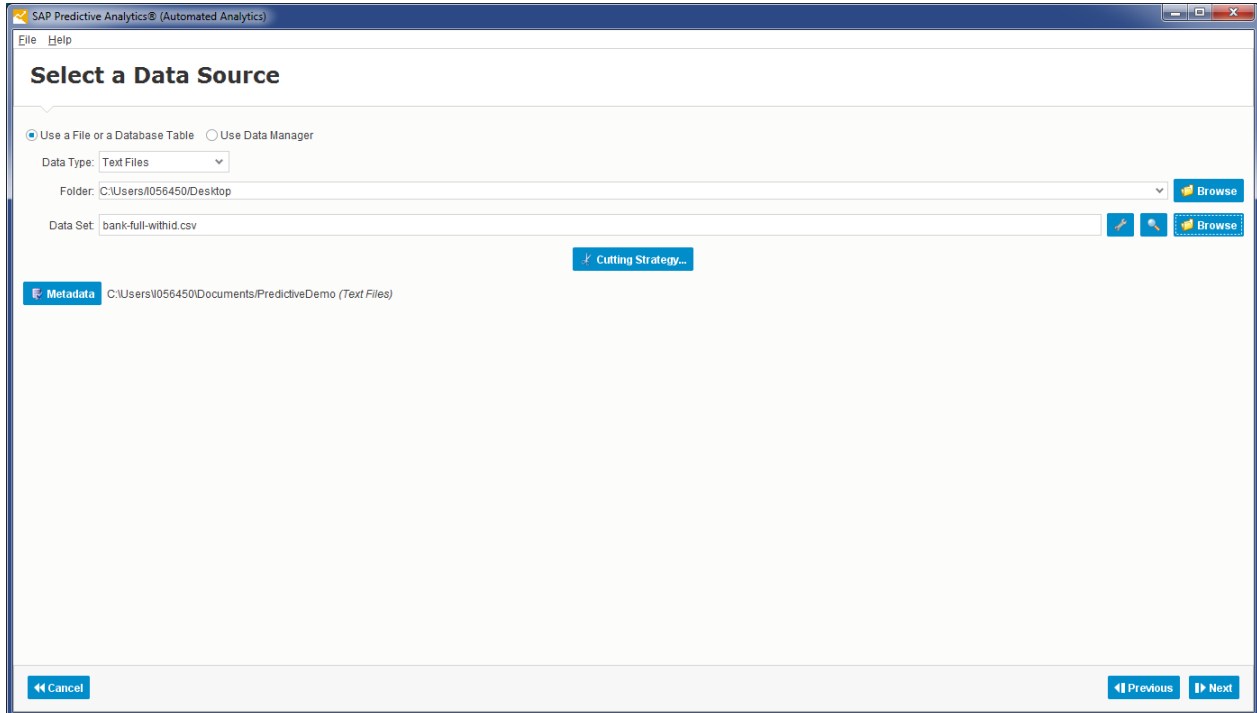




## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

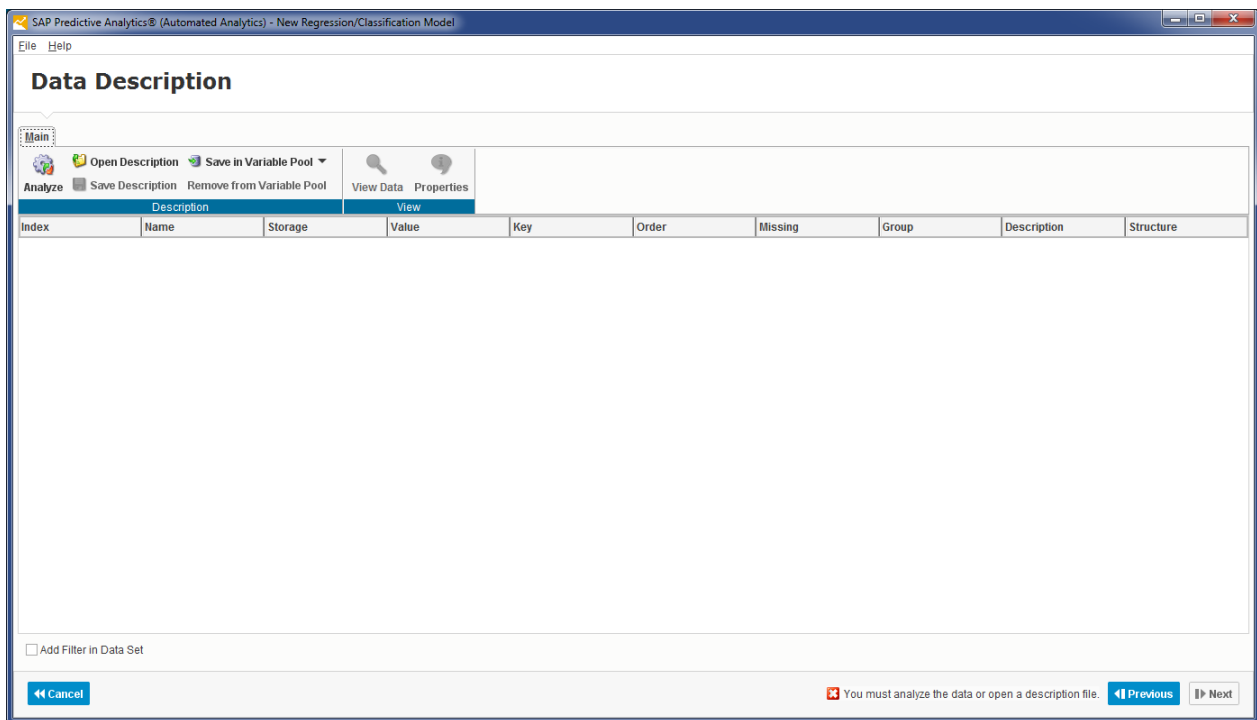
Now select the data source. Set the “Data Type” drop-down to “Text Files” and click the first “Browse” button. Here you specify the folder you saved the file bank-full-withid.csv into.

Select the second “Browse” button. Now select the file bank-full-withid.csv itself.

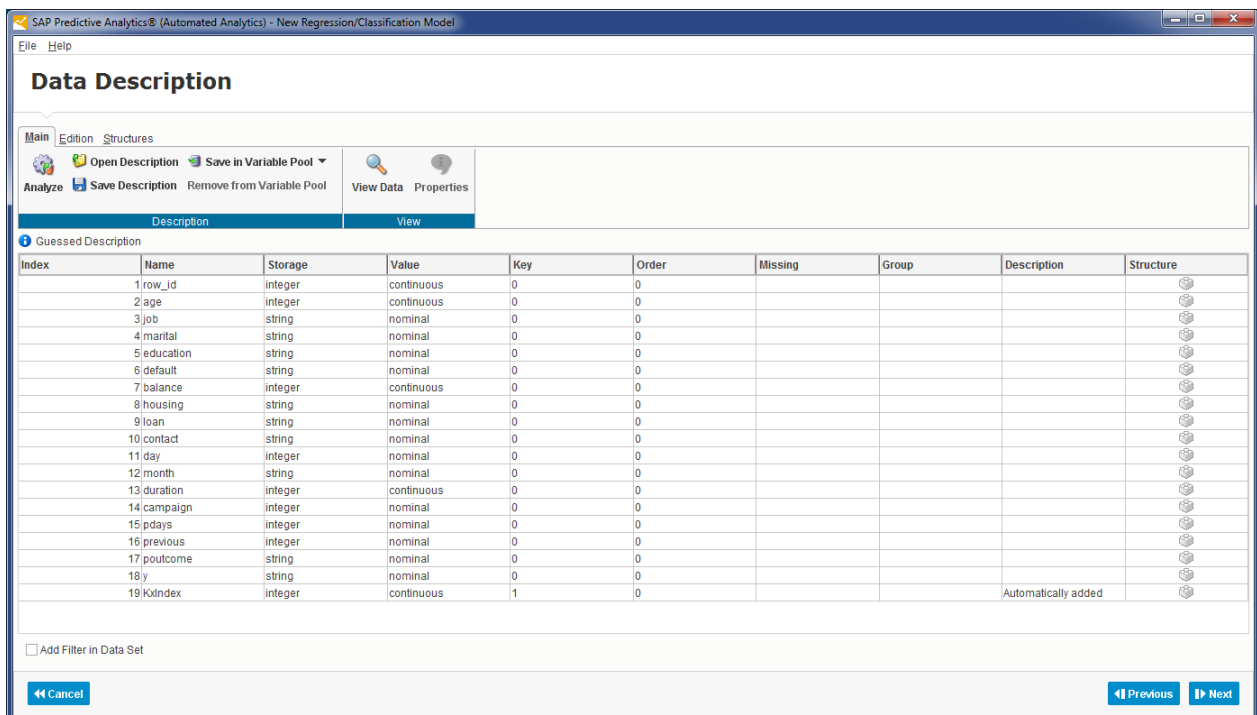


## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

For now, we are done specifying the data source and do not touch any of the other options. Click “Next”. You should now see the „Data Description” window.



Click the “Analyze” button so that SAP Predictive Analytics can determine the different columns, their name and data type. You could change some of the settings. However, for this tutorial we do not modify anything and keep the suggestions.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Click „View Data“ and you see a preview of the first rows from our dataset.

The screenshot shows the 'Automated Analytics Sample Data View' window. At the top, the 'Data Set' is 'bank-full-withid.csv'. Below it, 'First Row Index' is set to 1 and 'Last Row Index' is set to 100. The window has three tabs: 'Data' (selected), 'Statistics', and 'Graph'. The 'Data' tab displays a table with 12 columns: row\_id, age, job, marital, education, default, balance, housing, loan, contact, and day. The table contains 21 rows of data. At the bottom, there are icons for 'Data', 'Statistics', and 'Graph', a 'Help' button, and a 'Close' button.

|    | row_id | age | job          | marital  | education | default | balance | housing | loan | contact | day |
|----|--------|-----|--------------|----------|-----------|---------|---------|---------|------|---------|-----|
| 1  | 1      | 58  | managem...   | married  | tertiary  | no      | 2143    | yes     | no   | unknown |     |
| 2  | 2      | 44  | technician   | single   | secondary | no      | 29      | yes     | no   | unknown |     |
| 3  | 3      | 33  | entrepreneur | married  | secondary | no      | 2       | yes     | yes  | unknown |     |
| 4  | 4      | 47  | blue-collar  | married  | unknown   | no      | 1506    | yes     | no   | unknown |     |
| 5  | 5      | 33  | unknown      | single   | unknown   | no      | 1       | no      | no   | unknown |     |
| 6  | 6      | 35  | managem...   | married  | tertiary  | no      | 231     | yes     | no   | unknown |     |
| 7  | 7      | 28  | managem...   | single   | tertiary  | no      | 447     | yes     | yes  | unknown |     |
| 8  | 8      | 42  | entrepreneur | divorced | tertiary  | yes     | 2       | yes     | no   | unknown |     |
| 9  | 9      | 58  | retired      | married  | primary   | no      | 121     | yes     | no   | unknown |     |
| 10 | 10     | 43  | technician   | single   | secondary | no      | 593     | yes     | no   | unknown |     |
| 11 | 11     | 41  | admin.       | divorced | secondary | no      | 270     | yes     | no   | unknown |     |
| 12 | 12     | 29  | admin.       | single   | secondary | no      | 390     | yes     | no   | unknown |     |
| 13 | 13     | 53  | technician   | married  | secondary | no      | 6       | yes     | no   | unknown |     |
| 14 | 14     | 58  | technician   | married  | unknown   | no      | 71      | yes     | no   | unknown |     |
| 15 | 15     | 57  | services     | married  | secondary | no      | 162     | yes     | no   | unknown |     |
| 16 | 16     | 51  | retired      | married  | primary   | no      | 229     | yes     | no   | unknown |     |
| 17 | 17     | 45  | admin.       | single   | unknown   | no      | 13      | yes     | no   | unknown |     |
| 18 | 18     | 57  | blue-collar  | married  | primary   | no      | 52      | yes     | no   | unknown |     |
| 19 | 19     | 60  | retired      | married  | primary   | no      | 60      | yes     | no   | unknown |     |
| 20 | 20     | 33  | services     | married  | secondary | no      | 0       | yes     | no   | unknown |     |
| 21 | 21     | 28  | blue-collar  | married  | secondary | no      | 723     | yes     | yes  | unknown |     |

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Close this preview and hit “Next”.

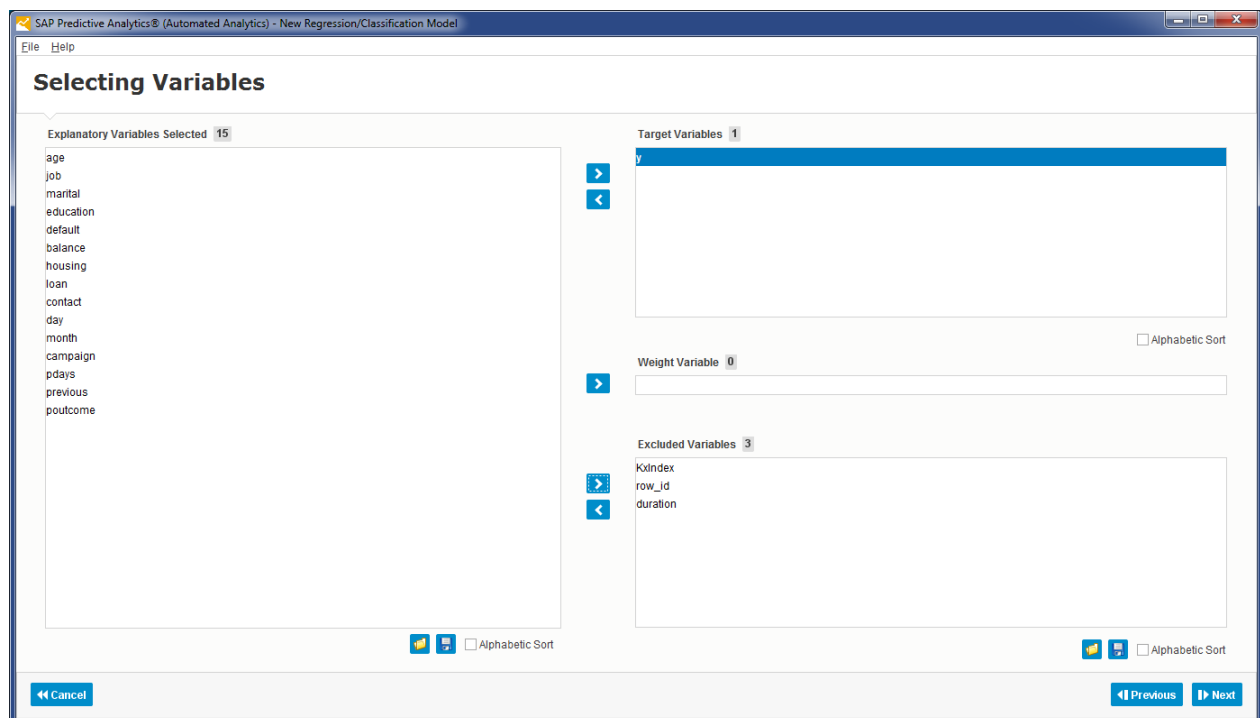
Now define the classification you are looking for and what data is available for the predictive model.

**Target Variables:** The binary variable our model should describe. SAP Predictive Analytics automatically selects the last column of the dataset as target variable, which is correct in this case. The column “y” indicates whether a customer did or did not sign up for a term deposit.

**Weight Variable:** Some datasets include a weight variable to create an analysis that is representative of a larger population. This option is not required for our dataset.

**Excluded Variables:** Any variables that should be ignored when building the model. Select the variables “row\_id” and “duration” to be excluded. The duration states how long a person, that had already been called, stayed on the phone for. Since we want to determine who to call, we do not have that information available for the prediction phase. Hence the column needs to be eliminated.

Notice that a variable called “KxIndex” has also been excluded by default. This is a variable created internally by SAP Predictive Analytics.

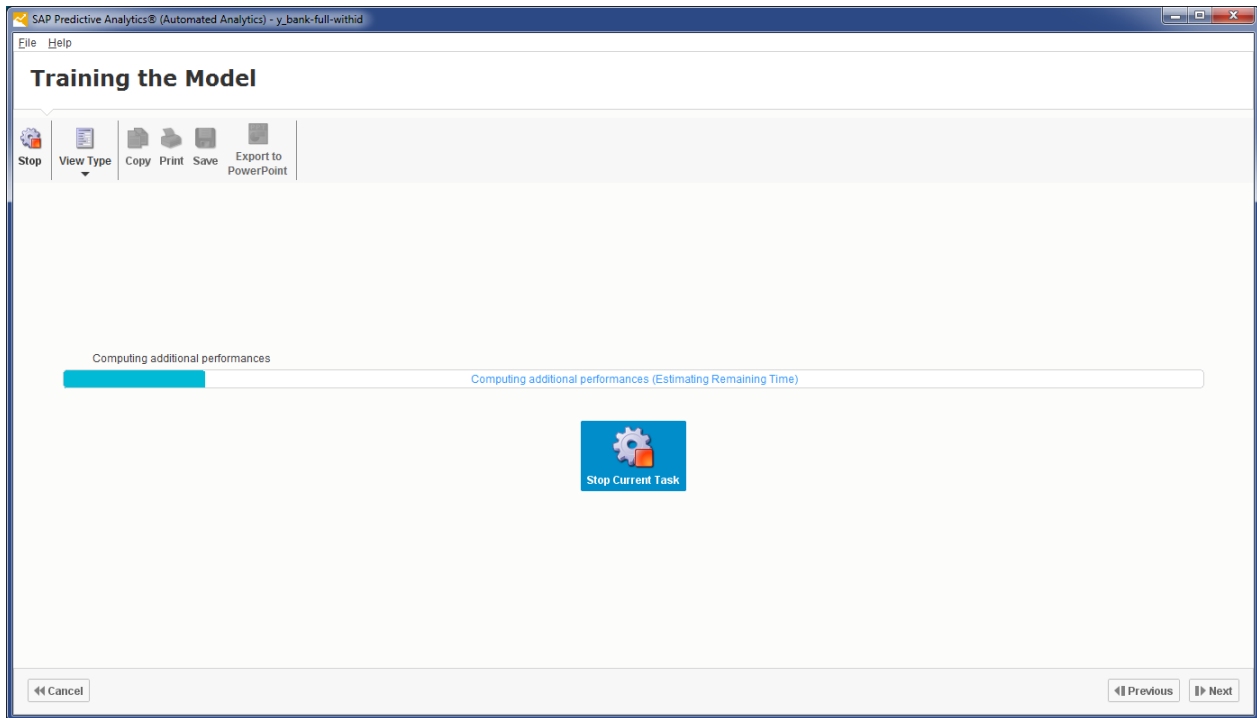


## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Click “Next”. Then hit “Generate” to create the model. This is where the model is automatically created. A very comprehensive framework, that does not make any assumptions on the data, creates the predictive model.

Due to the high level of automation such models can be mass produced. The models however are not a black box. They are described in a lot of detail giving the user confidence in its workings.

If you would like to better understand what happens in this seemingly magic phase, see the footnote on page 3 for an article that explains how SAP Predictive Analytics can find the best model.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Once the model is created, you will see this summary.

The most important measures are:

**Predictive Power (KI):** Indicates the quality of the model. The value range is between 0 and 1, where 0 indicates a random result and a value of 1 would be the perfect model.

**Prediction Confidence (KR):** Indicates the robustness of the model. It describes the ability of the model to achieve the same performance when it is applied to a new dataset. The range is also between 0 and 1. Typically you want this value to be larger than 0.95.

With a Predictive Power of 0.5130 and a Prediction Confidence of 0.9912 we can be happy with the model.

You can also see that out of the 15 input variables only 14 have been used.

Hint: SAP Predictive Analytics will build the model towards target variable's less common value. In our data only 11.75% of the customers did sign up for the term deposit. Hence the model describes the customers that decided to buy.

The screenshot displays the SAP Predictive Analytics (Automated Analytics) - y\_bank-full-withid window. The title bar indicates the application name. The main window has a menu bar with 'File' and 'Help'. Below the menu bar is a toolbar with icons for 'Stop', 'View Type', 'Copy', 'Print', 'Save', and 'Export to PowerPoint'. The 'View Type' dropdown is set to 'Model Overview'. The main content area is titled 'Overview' and shows the following information:

**Model: y\_bank-full-withid**

|                               |                       |
|-------------------------------|-----------------------|
| Data Set:                     | bank-full-withid.csv  |
| Initial Number of Variables:  | 19                    |
| Number of Selected Variables: | 15                    |
| Number of Records:            | 45,211                |
| Building Date:                | 2015-08-08 12:49:43   |
| Learning Time:                | 10s                   |
| Engine Name:                  | Kxen.RobustRegression |
| Author:                       | I056450               |

**Monotonic Variables**

| Variable | Value      | Storage | Role  | Monotonicity |
|----------|------------|---------|-------|--------------|
| age      | continuous | integer | input | decrease     |

**Nominal Targets**

| Target Key      | yes    |
|-----------------|--------|
| no - Frequency  | 88.27% |
| yes - Frequency | 11.73% |

**Selection Process Selected Iteration**

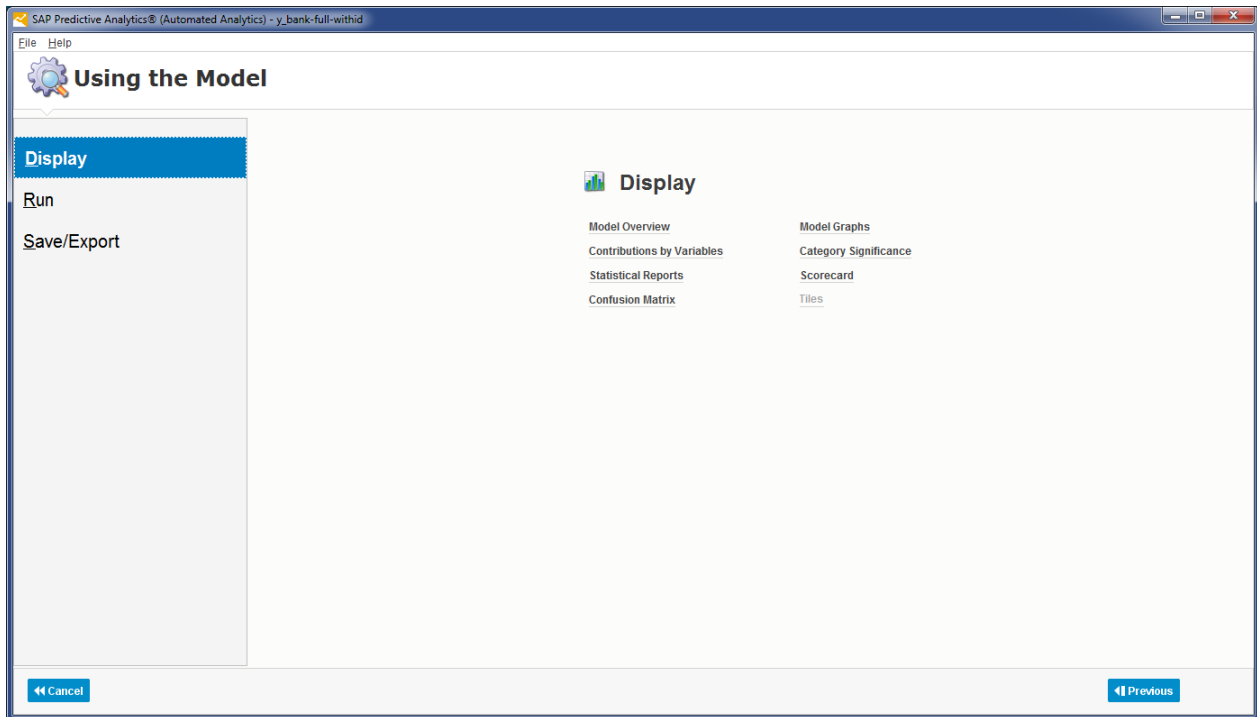
| Iteration | Predictive Power (KI) | Prediction Confidence (KR) | Nb. Variables Kept |
|-----------|-----------------------|----------------------------|--------------------|
| 1         | 0.5130                | 0.9912                     | 14                 |

At the bottom of the window, there are buttons for 'Cancel', 'Previous', and 'Next'.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

### Understand Predictive Model

Now that the model is built we want to understand it before using it. Click “Next”, then “Model Overview”.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

At first you see the same description as before. However, change the drop down on top to “Executive Report” and you will see more details.

The screenshot shows the SAP Predictive Analytics (Automated Analytics) interface for a model named 'y\_bank-full-withid'. The 'Report Type' is set to 'Executive Report'. The 'Model Overview' section displays the following details:

| Model: y_bank-full-withid          |
|------------------------------------|
| Data Set: bank-full-withid.csv     |
| Initial Number of Variables: 19    |
| Number of Selected Variables: 15   |
| Number of Records: 45,211          |
| Building Date: 2015-08-08 12:49:43 |
| Learning Time: 10s                 |
| Engine Name: Kxen.RobustRegression |
| Author: I056450                    |

Below the overview, the 'Monotonic Variables' section shows a table with one variable:

| Variable | Value      | Storage | Role  | Monotonicity |
|----------|------------|---------|-------|--------------|
| age      | continuous | integer | Input | decrease     |

The 'Nominal Targets' section shows the target variable 'y' with the following distribution:

| Target Key      | yes    |
|-----------------|--------|
| no - Frequency  | 88.27% |
| yes - Frequency | 11.73% |

At the bottom of the window, there are 'Cancel' and 'Previous' buttons.



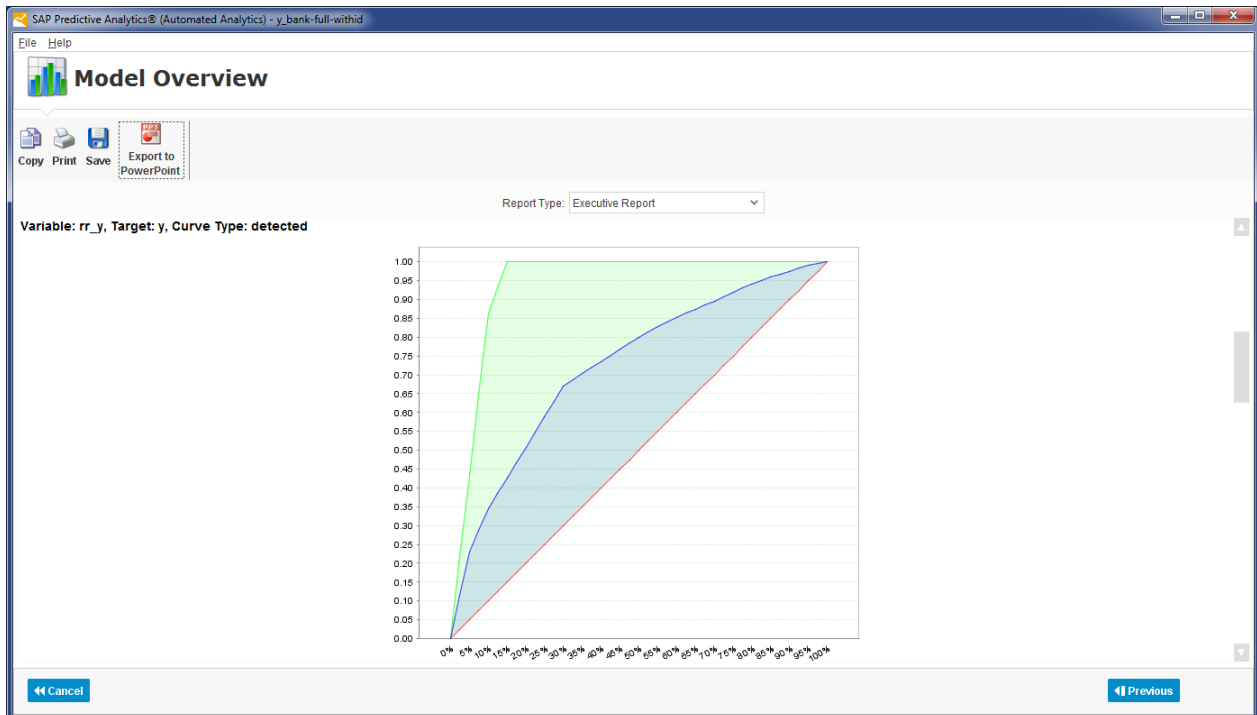
## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Scroll down to the Profit Curve, which shows the quality of the model. The x-axis shows the percentage of the total dataset. The y-axis shows the percentage of the correctly identified target.

The **red line** indicates the random distribution of our target variable. If you contact 10% of the customers, you have identified 10% of the target. Hence the line is linear.

The **green line** indicates the perfect model. All targets are identified first.

The **blue line** is our model. The closer it is to the top left corner the better is the model.

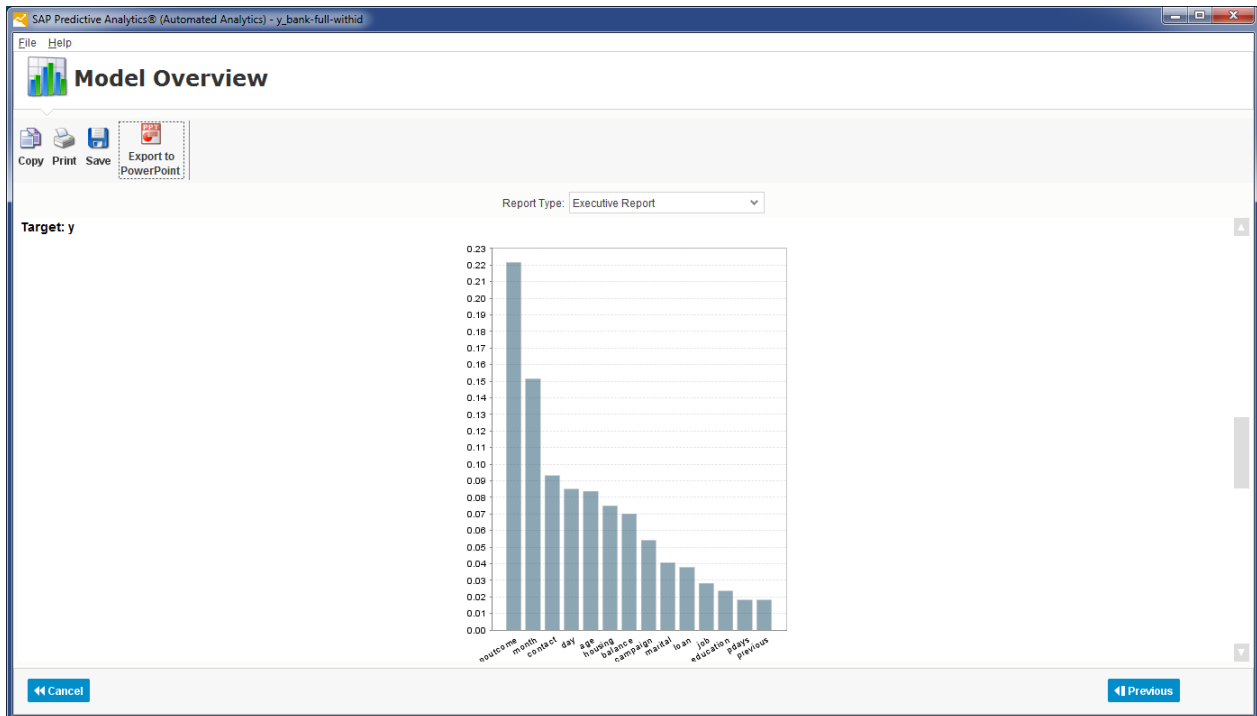


## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Scroll down further and you find the Contributions Chart. It explains which input parameters were selected and which ones had the strongest influence on our target variable. The chart needs to be read from left to right.

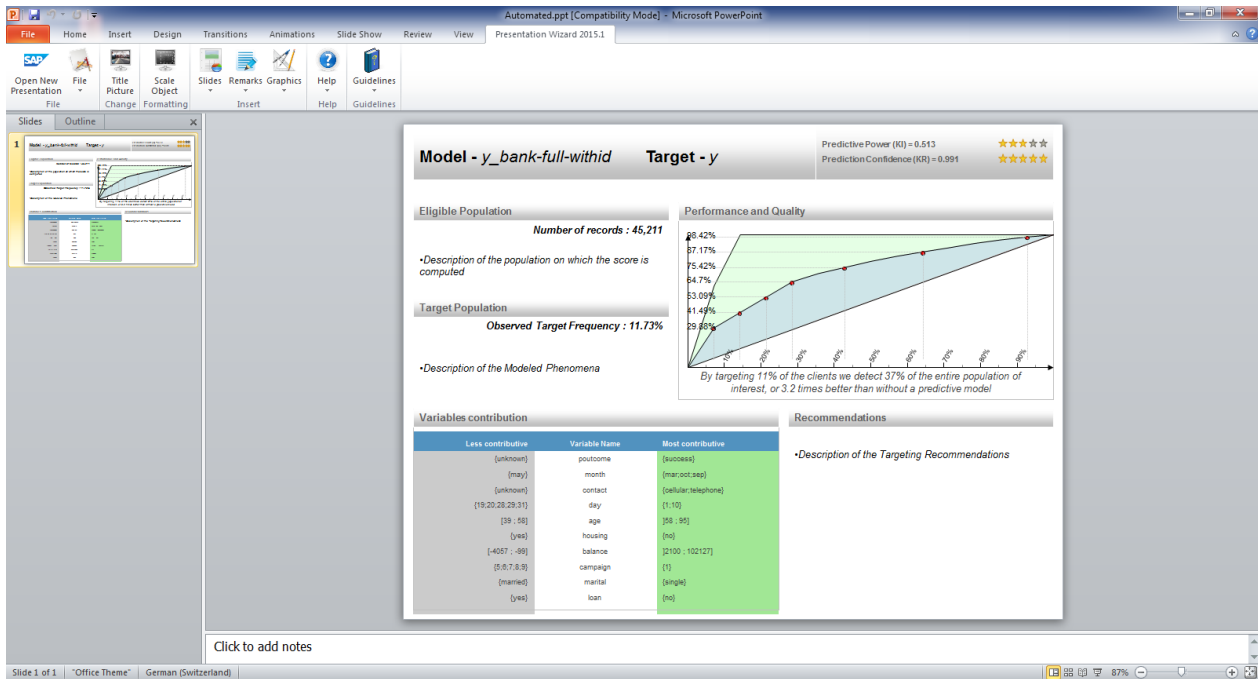
You see that the column “outcome” has by far the strongest impact followed by “month” and “contact”. So the most important information to understand whether the customer will subscribe to a term deposit is the outcome of the previous marketing campaign.

This page is just giving an overview. A few clicks later we will drill into the columns to see exact impact, for instance which months had the strongest impact on whether the customer did subscribe or not.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

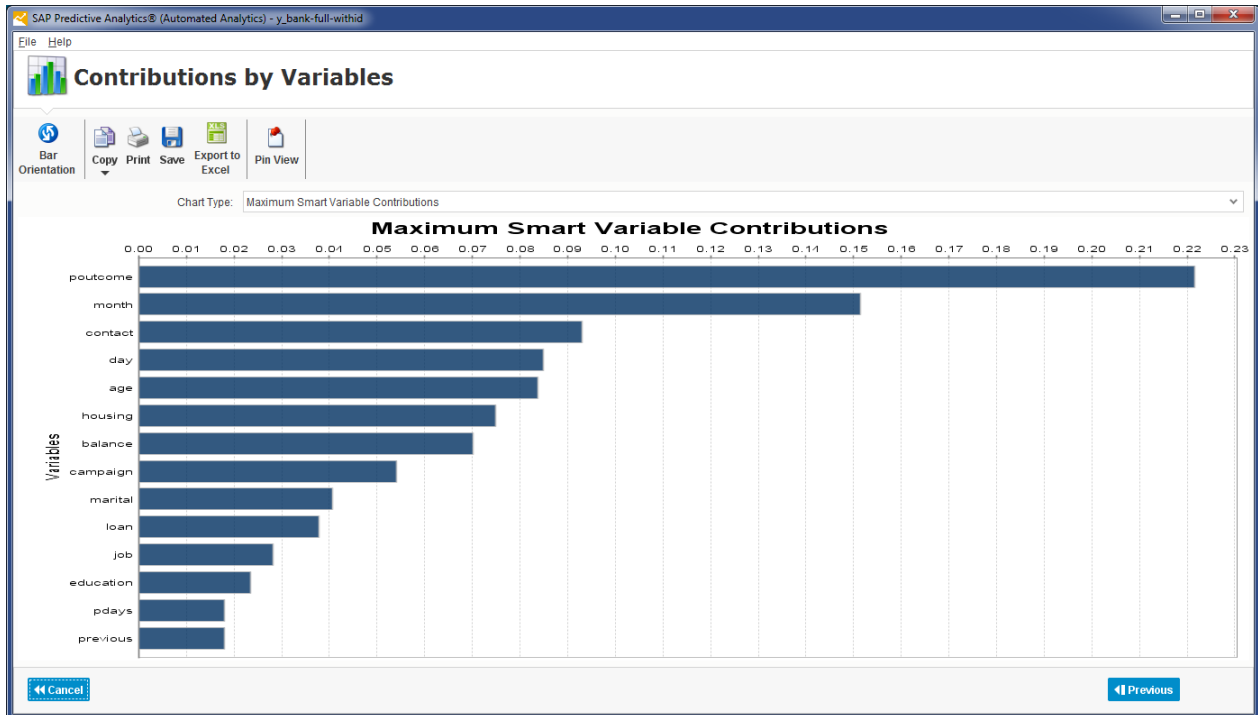
To document the model just click the “Export to Powerpoint” icon in the toolbar. Give the presentation file a name and open it MS Powerpoint. The model is documented and could be used in a presentation immediately.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Now we want to better understand the impact of the variables on the outcome of the Marketing campaign. Go back to SAP Predictive Analytics, click “Previous” and select “Contributions by Variables”.

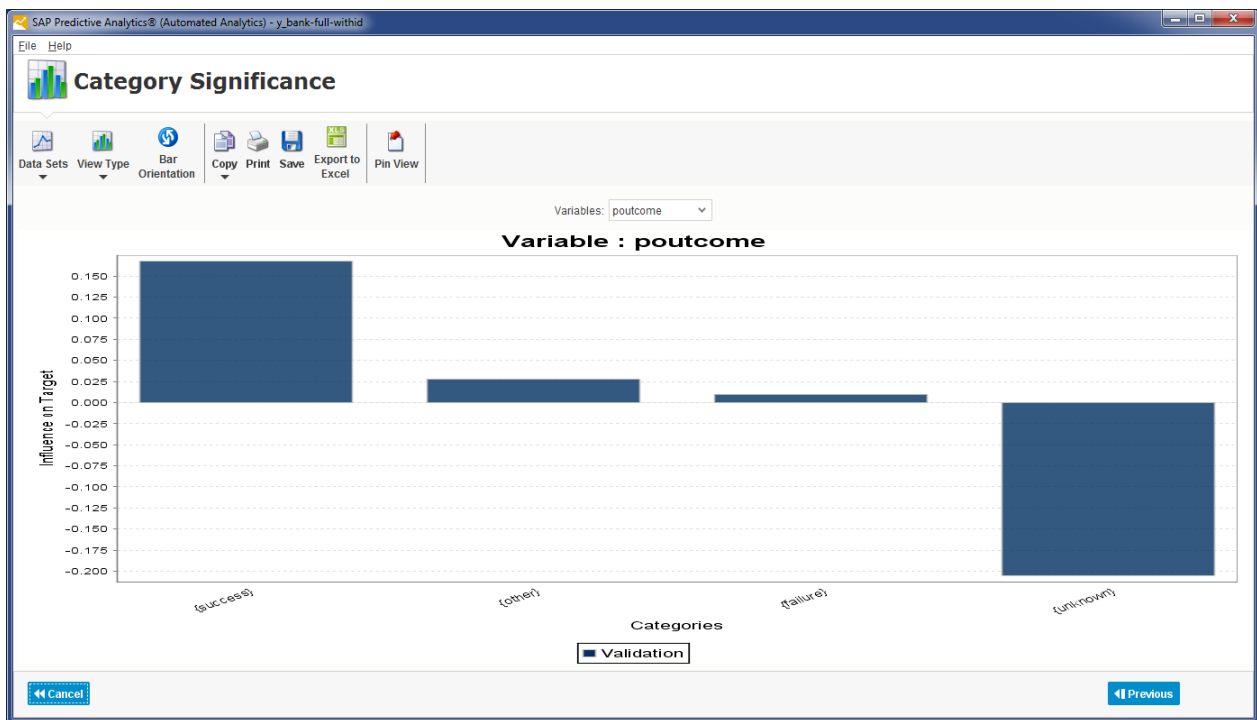
You should now see the same Contributions Chart as before. This chart, however, is interactive.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Double-click the “poutcome” bar and you can understand how the content of this variable impacts the target variable. This chart has to be read from left to right. The bar on the very left has had the strongest positive impact on the target variable. On the right hand-side you see the strongest negative impact on the target variable.

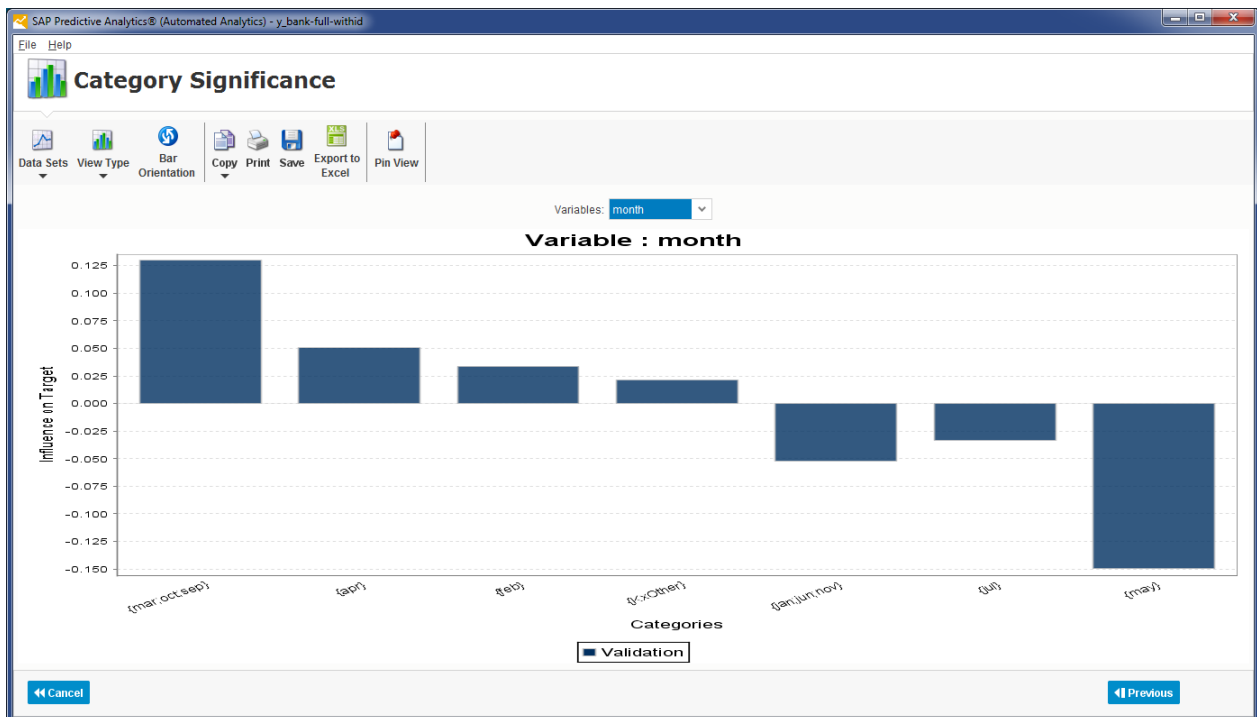
You see that if “poutcome” has the value “success”, then the probability of a sale is strongest. So this means, the most important contributing factor to whether the customer will respond positively in this campaign or not, is whether the customer did sign up in the previous campaign as well.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

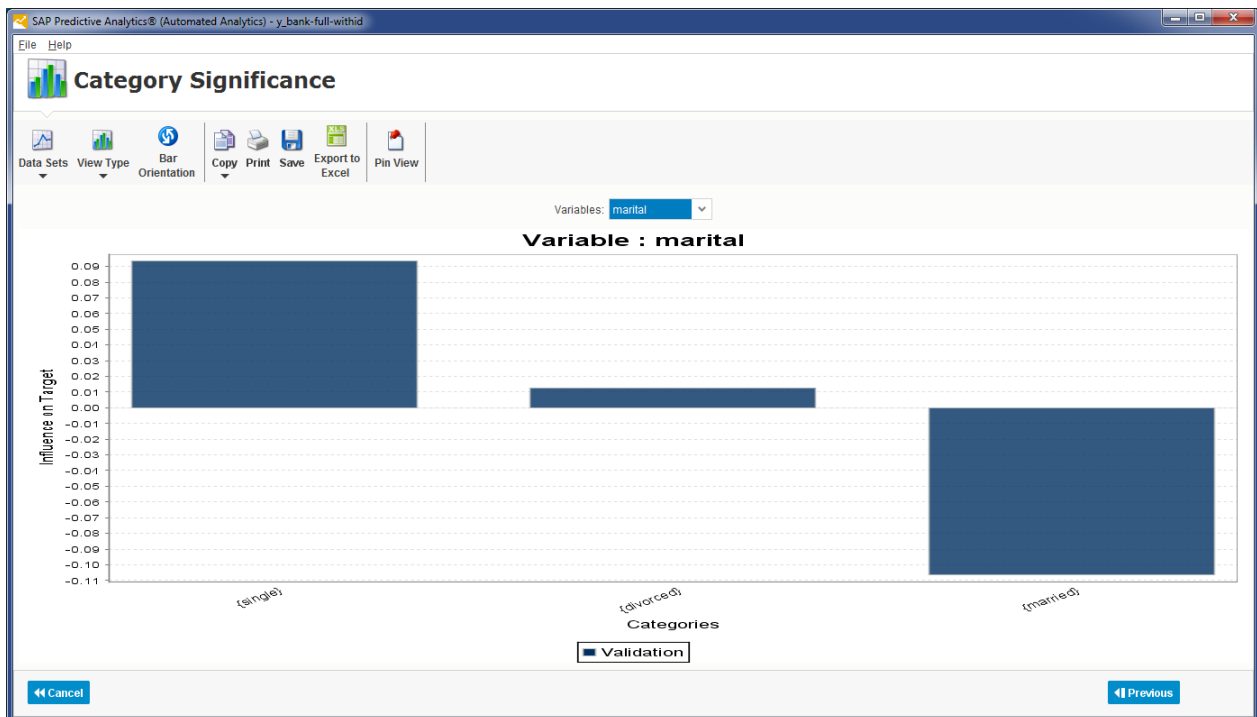
Set the drop-down above the chart to “month”. You see that the campaign was most successful in March, September and October. Customers showed similar behavior in these months towards the target variable, therefore SAP Predictive Analytics automatically grouped the values to increase the model’s robustness.

May however had the most negative impact. This could be due to our customers needing their money for the summer holidays.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Set the drop-down to “marital” and you see that the impact on the target is strongest if the person is single.



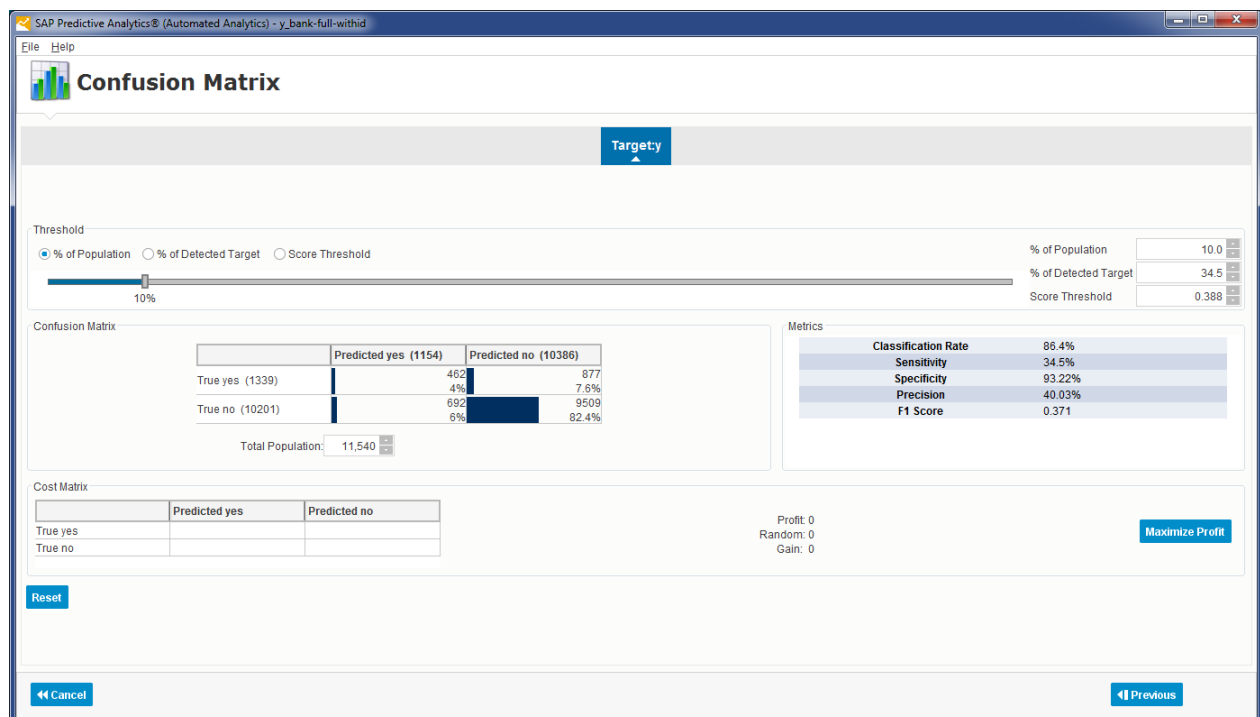
## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Now that we understand the model, let's have a look at the value this model is adding. Go back to the main screen. Most likely you have to hit "Previous" twice, and select the "Confusion Matrix".

The confusion matrix tells us how well the model predicts the correct outcome. Imagine you have budget to contact 10% of your contacts. If you contact your customers randomly, you will only reach 10% of the customers that purchase. However, with a good predictive model you will be able to increase the success rate.

Set the value for "% of Population" to 10.0%. The "Percentage of Detected Target" is shown as 34.5%!

This means, that the model increases the efficiency of your Marketing campaign by the factor of 3.5 compared to selecting the recipients randomly. The model we created clearly improves the success rate of the marketing campaign.

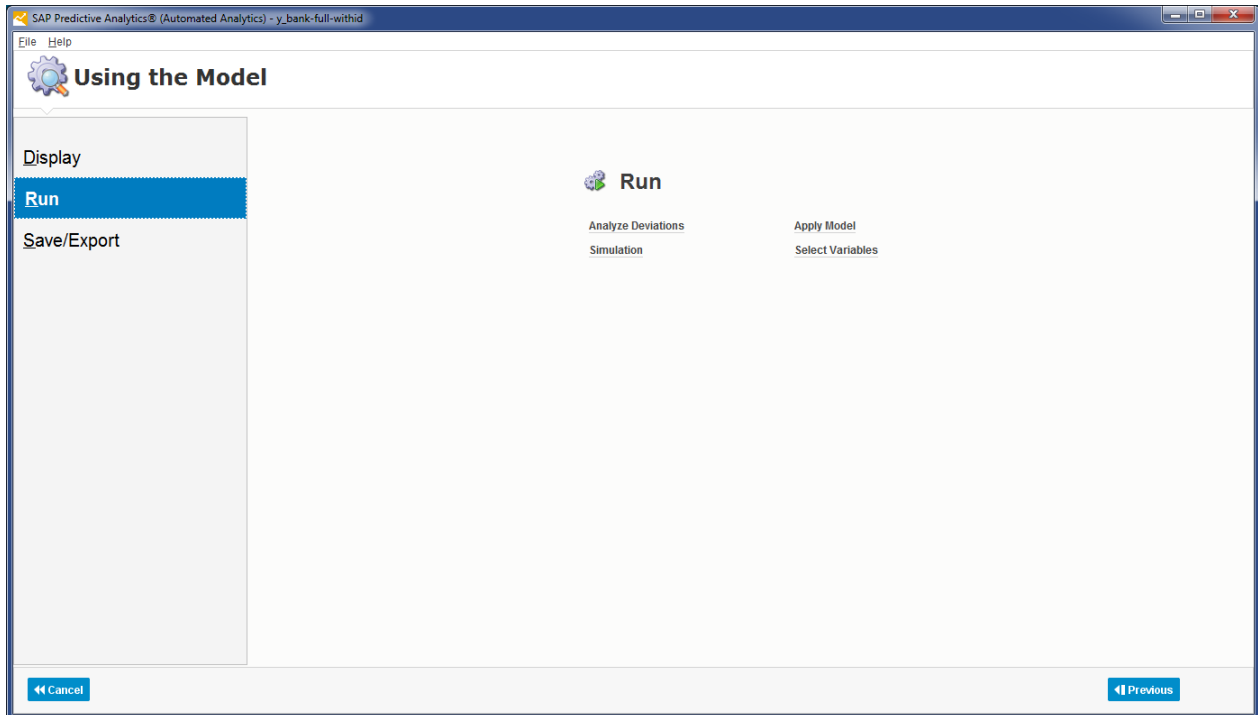




## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

### Apply Predictive Model

Now that we understand the model and we are convinced of its value, we want to put it into practice. Select “Previous” and click into the “Run” section.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Now we want to use the model to score customers we have not called yet on their likelihood to purchase the term deposit. For simplicity of this guide we will apply the model on the same dataset that we created the model on.

Select “Apply Model” and set the “Application Data Set” to the same file as before (bank-full-withid.csv).

We want to calculate the probability of the classification being “yes”. Change the “Generate” option to “Probability & Error Bars”.

Have the scores written to a flat file in the same folder. Call it “term-deposit-probability.csv”.

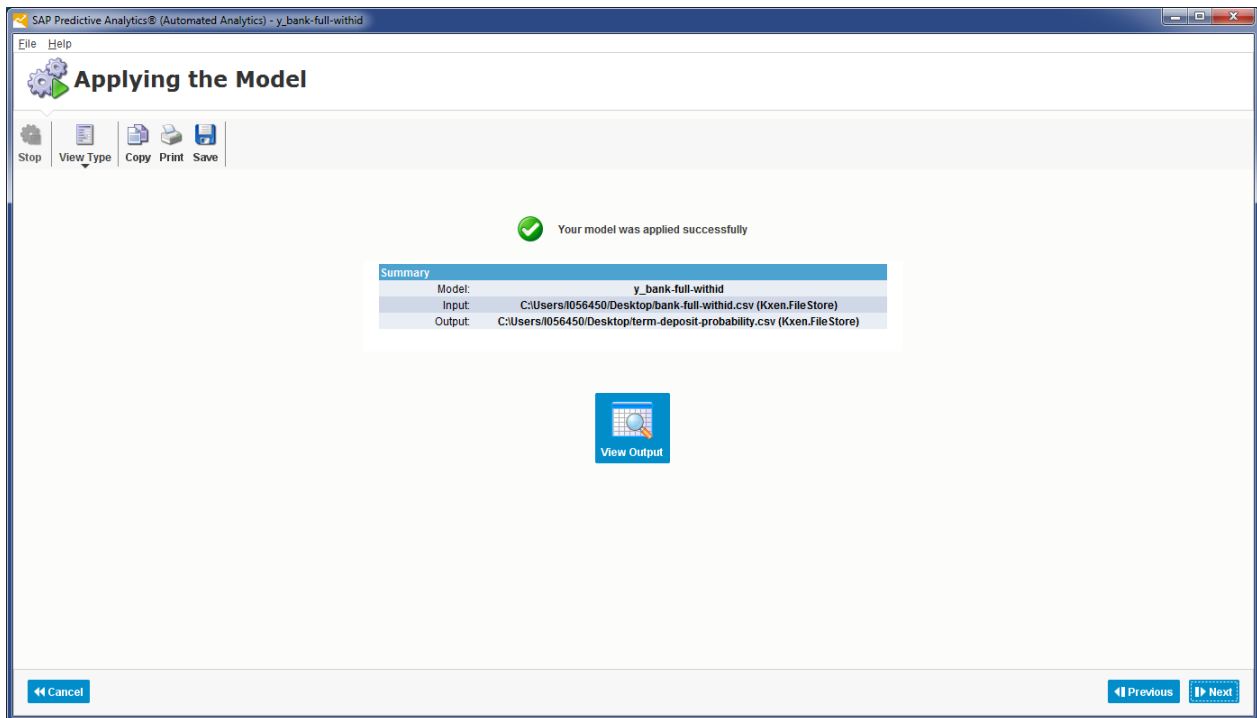
The screenshot shows the 'Applying the Model' dialog in the SAP Predictive Analytics (Automated Analytics) - y\_bank-full-withid window. The dialog is divided into several sections:

- Application Data Set:**
  - Data Type: Text Files (dropdown)
  - Folder: C:\Users\I056450\Desktop (text field with a dropdown arrow and a 'Browse' button)
  - Data: bank-full-withid.csv (text field with a 'Browse' button)
  - Buttons: Define Mapping
- Generation Options:**
  - Generate: Probability & Error Bars (dropdown with an 'Advanced Apply Settings...' button)
  - Mode: Apply (dropdown)
  - ☐ Add Score Deviation
- Results Generated by the Model:**
  - Data Type: Text Files (dropdown)
  - Folder: C:\Users\I056450\Desktop (text field with a dropdown arrow and a 'Browse' button)
  - Data: term-deposit-probability.csv (text field with a 'Browse' button)
  - Buttons: Define Mapping
- ☐ Use Direct Apply in the Database

At the bottom of the dialog, there are three buttons: Cancel, Previous, and Apply.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Click “Apply”. Once the model has been applied, you can see the results by clicking on “View Output” or by opening the file “term-deposit-probability.csv”.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

The probability score whether the customer will purchase is stored in the “proba\_rr\_y” column. You could sort the dataset on this column and pass the top 10% to the Marketing department as target group for the campaign.

**SAP Predictive Analytics (Automated Analytics) - y\_bank-full-withid**

File Help

### Applying the Model

Stop View Type Copy Print Save

Output Name: term-deposit-probability.csv

First Row Index: 1 Last Row Index: 100

**Data** Statistics Graph

|    | KxIndex | y     | rr_y          | proba_rr_y   | bar_rr_y     |
|----|---------|-------|---------------|--------------|--------------|
| 1  |         | 1 no  | 0.1009784...  | 0.0533159... | 0.8596448... |
| 2  |         | 2 no  | -0.0184235... | 0.0363924... | 0.5494147... |
| 3  |         | 3 no  | -0.1002477... | 0.0245768... | 0.4399189... |
| 4  |         | 4 no  | 0.0000819...  | 0.0502533... | 0.7891909... |
| 5  |         | 5 no  | 0.0823688...  | 0.0530565... | 0.7989225... |
| 6  |         | 6 no  | 0.0111836...  | 0.0510282... | 0.6277115... |
| 7  |         | 7 no  | 0.0802428...  | 0.0530289... | 0.7961206... |
| 8  |         | 8 no  | -0.0377530... | 0.0350729... | 0.4841339... |
| 9  |         | 9 no  | 0.0345459...  | 0.0524307... | 0.7140146... |
| 10 |         | 10 no | 0.0262384...  | 0.0520847... | 0.7254754... |
| 11 |         | 11 no | -0.0102423... | 0.0372554... | 0.5935362... |
| 12 |         | 12 no | 0.0997649...  | 0.0532990... | 0.8524955... |
| 13 |         | 13 no | -0.0704004... | 0.0337466... | 0.5548710... |
| 14 |         | 14 no | 0.0255734...  | 0.0520380... | 0.7293627... |
| 15 |         | 15 no | -0.0498144... | 0.0345829... | 0.4920826... |
| 16 |         | 16 no | -0.0198064... | 0.0362809... | 0.5578334... |
| 17 |         | 17 no | -0.0052964... | 0.0430066... | 0.6536596... |
| 18 |         | 18 no | -0.0789602... | 0.0320390... | 0.5536185... |
| 19 |         | 19 no | 0.0848380...  | 0.0530909... | 0.8021765... |

Cancel Previous Next

### SUMMARY

You have successfully used SAP Predictive Analytics to optimize a Marketing campaign. The Marketing budget is spent on the customers with the highest chance of success, thereby significantly increasing the success rate of the next campaign.

No statistical expert knowledge was needed to create these results, which can be easily embedded into the existing processes of the business users.

Similarly, many different data mining tasks can be carried out with SAP Predictive Analytics.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

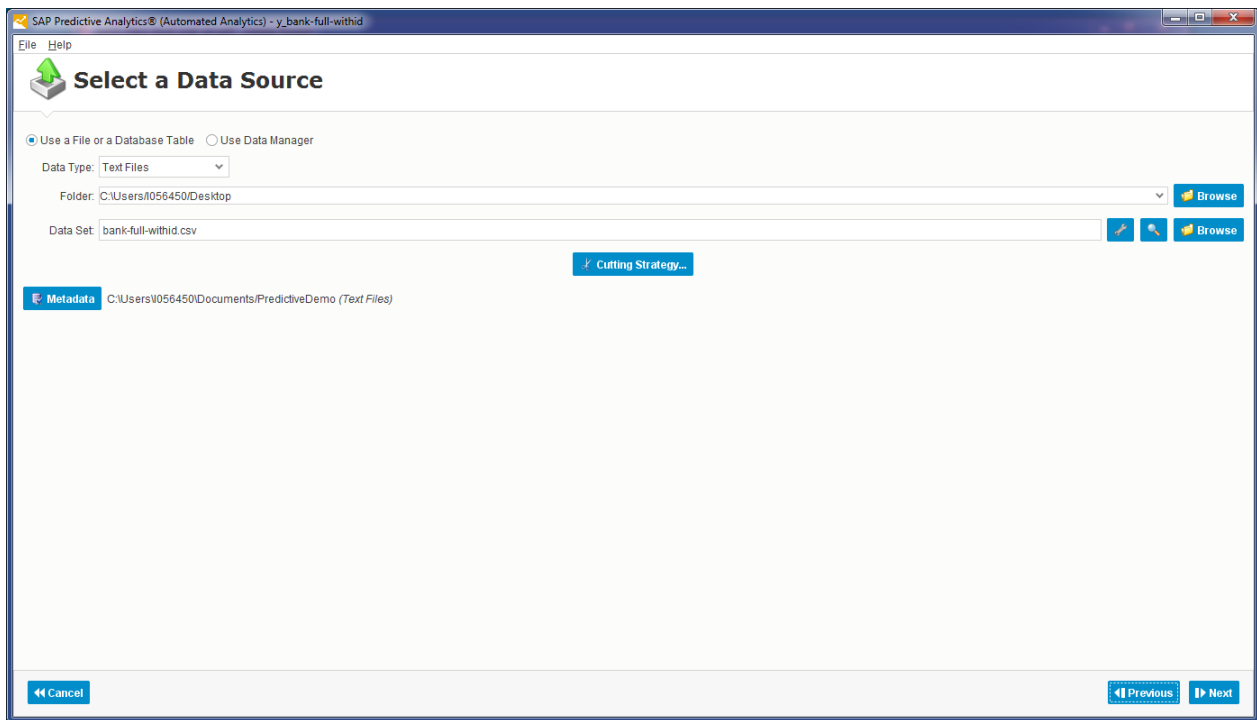
### OPTIONAL MODIFICATIONS

So far we have created a predictive model with just a few clicks, keeping most of the default settings. In these optional modifications we are looking at some additional aspects to extend the model or its usage.

#### Model Performance on new Data

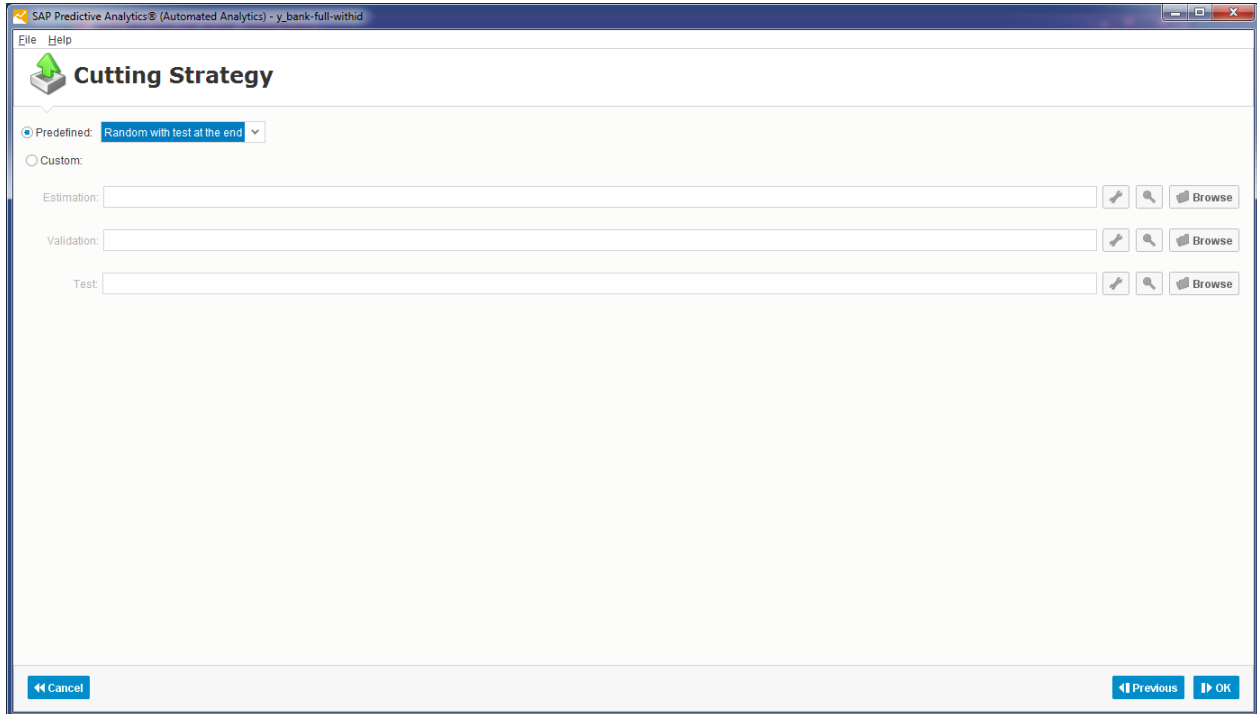
When creating a model with the default setting as done before, the data passed into SAP Predictive Analytics is split in two parts: to create new models and to select the best model.

If you would like to see the model's performance on completely new data you need to change the "Cutting Strategy". Start creating the predictive model as before until you get to this screen.



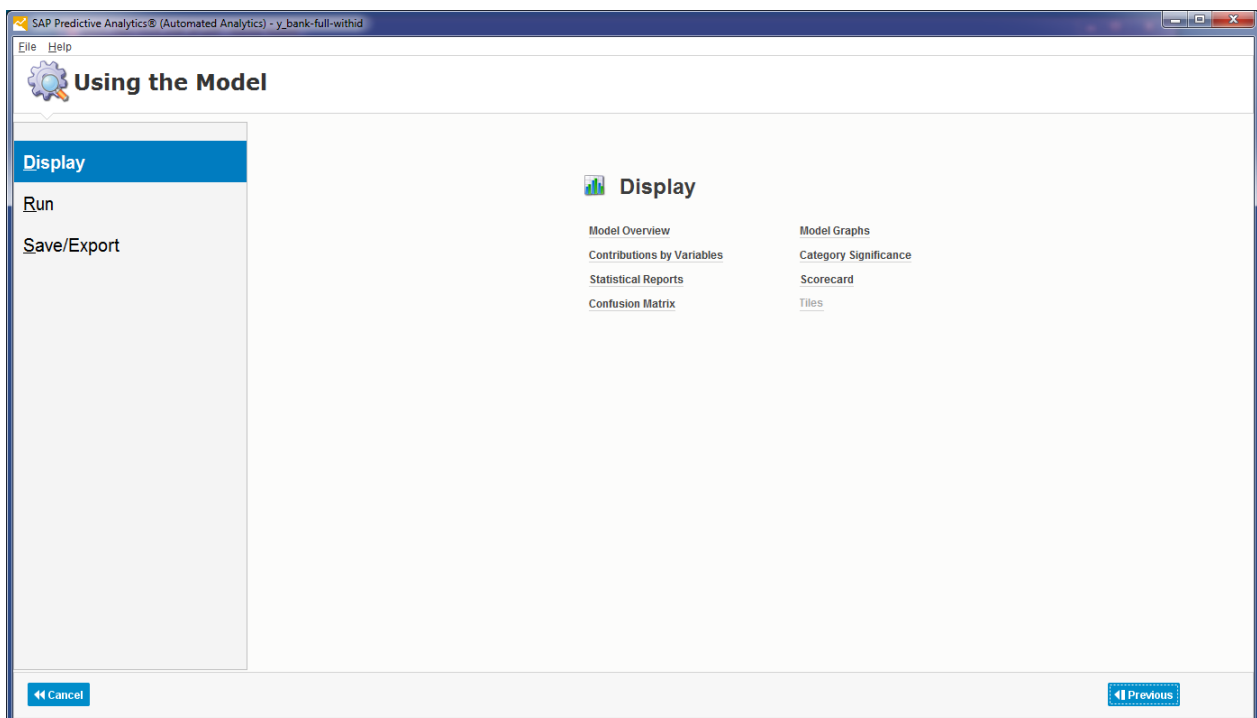
## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Click on “Cutting Strategy” and change the drop down to “Random with test at the end”. The data split now includes a part that is not used at all during the model creation. It will only be used at the very end of the process to describe the model's performance on previously unseen data.



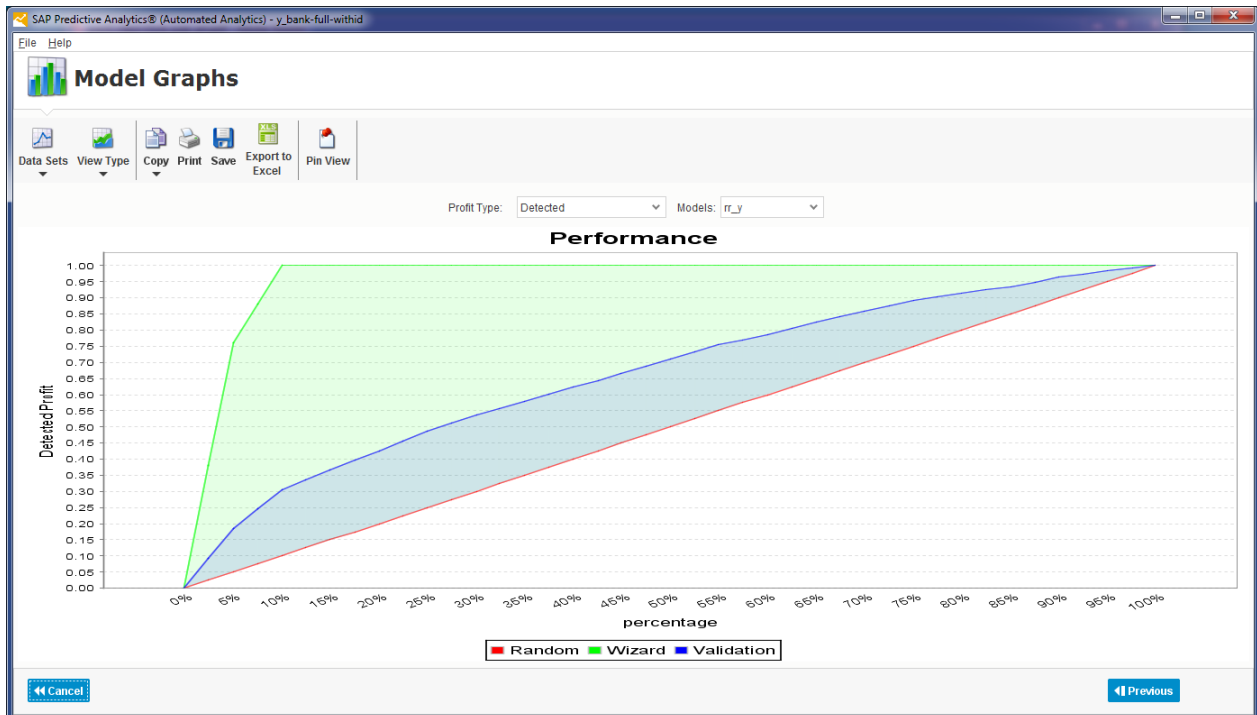
Then click “OK” and continue creating the model as before.

Don not worry for now in case you get a warning message about a deviation in the target variable y. Once the model has been created you get to this screen. Click “Model Graphs”.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

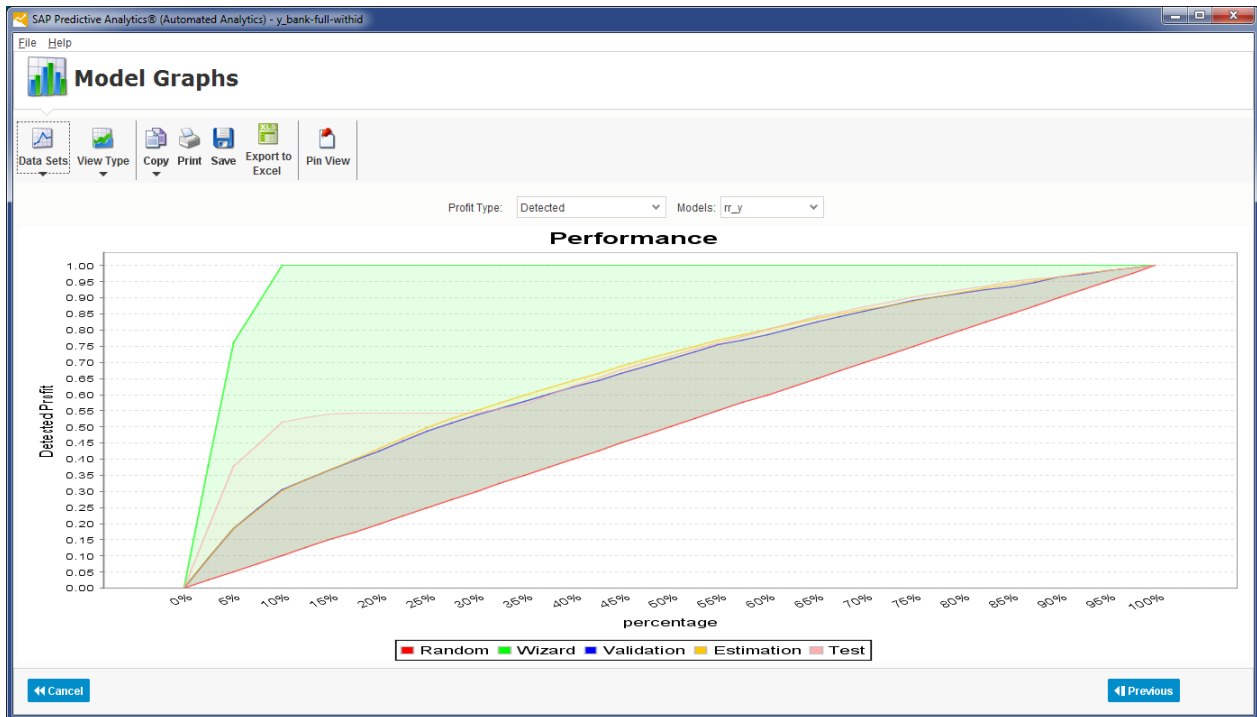
At first you will see the model's performance chart, similar to before.





## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Now click on “Data Sets” and “All Data Sets” to see the performance of the new data named “Test”.



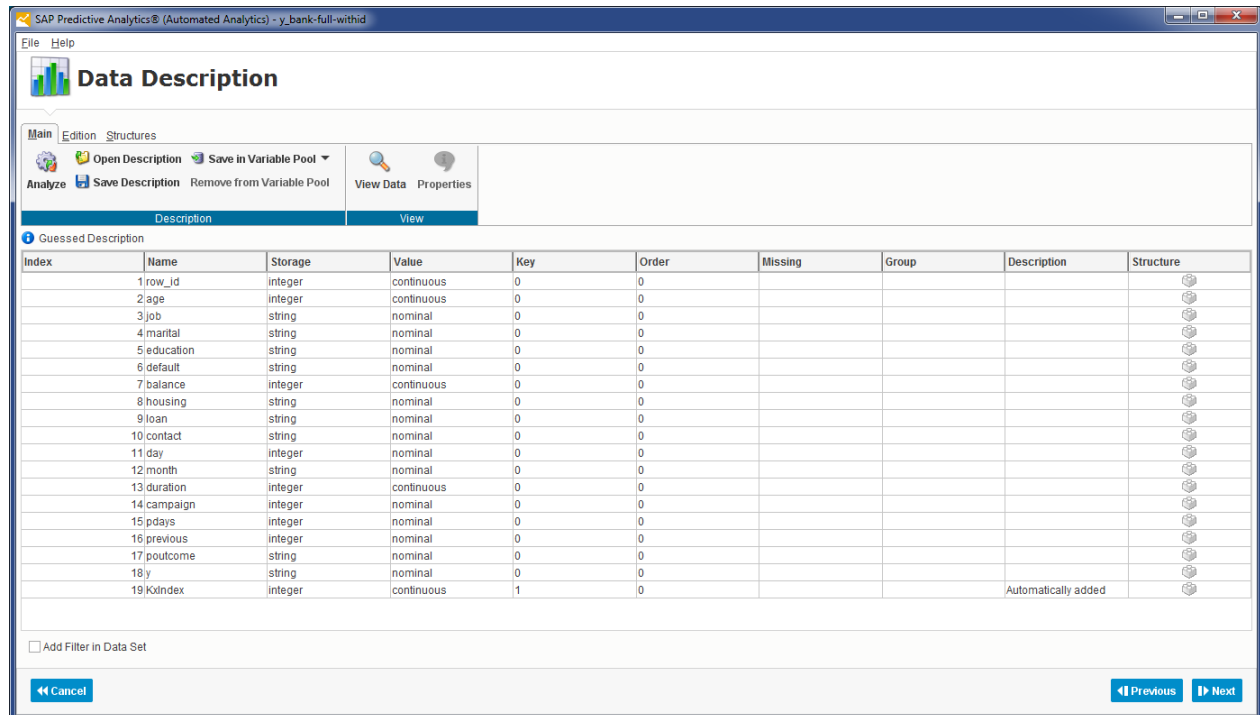
It turns out that the model performs better than expected on the top end of the calculated probabilities. You can see how steep the curve for the “Test” data rises before levelling off.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

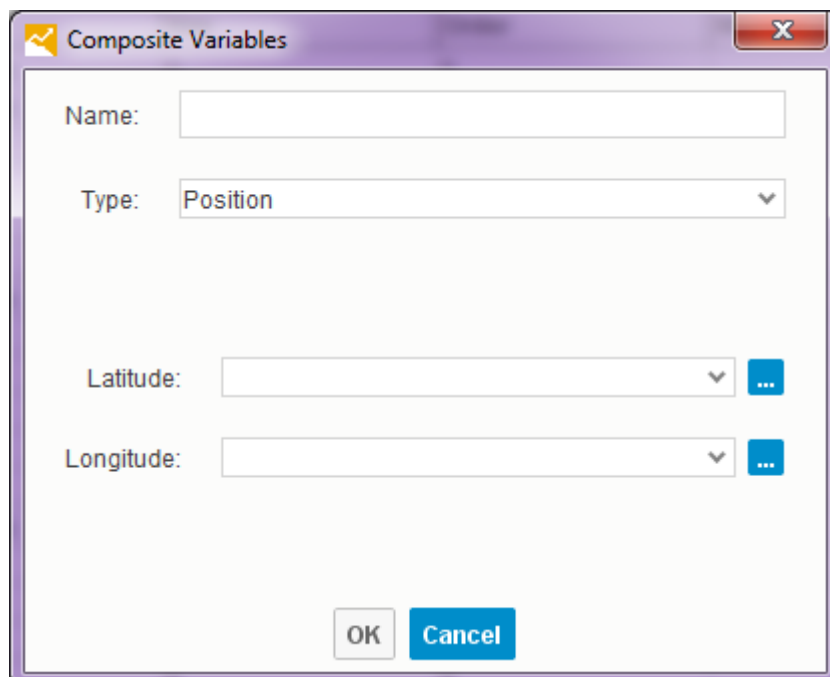
### Describe the customer in more detail

Often a model can be enhanced by combining multiple variables. SAP Predictive Analytics can create such composite variables even when connecting to a flat file.

Start creating a model as before until you get to this screen.



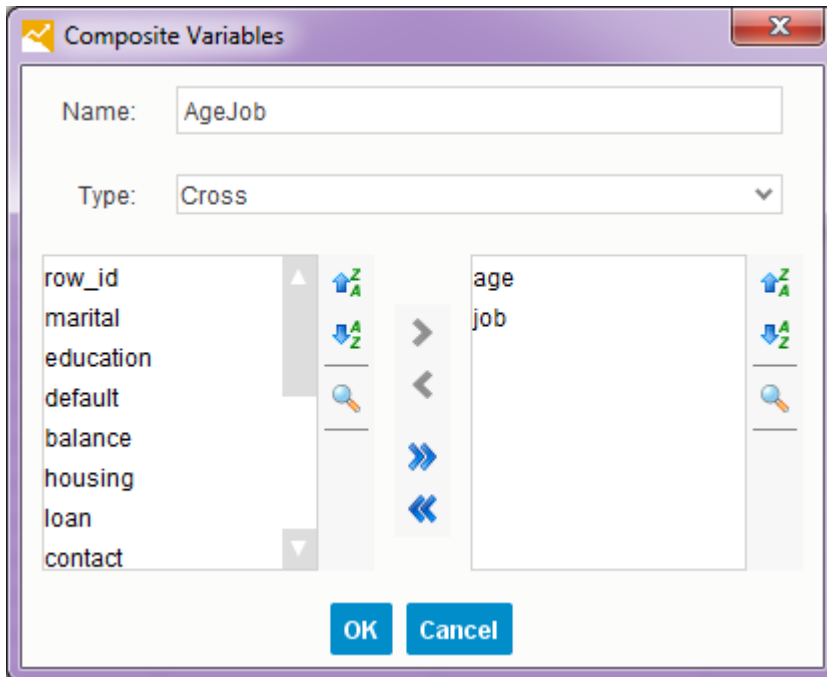
Click the "Edition" tab on top. Then select "Composite Variables" and click the plus-sign.



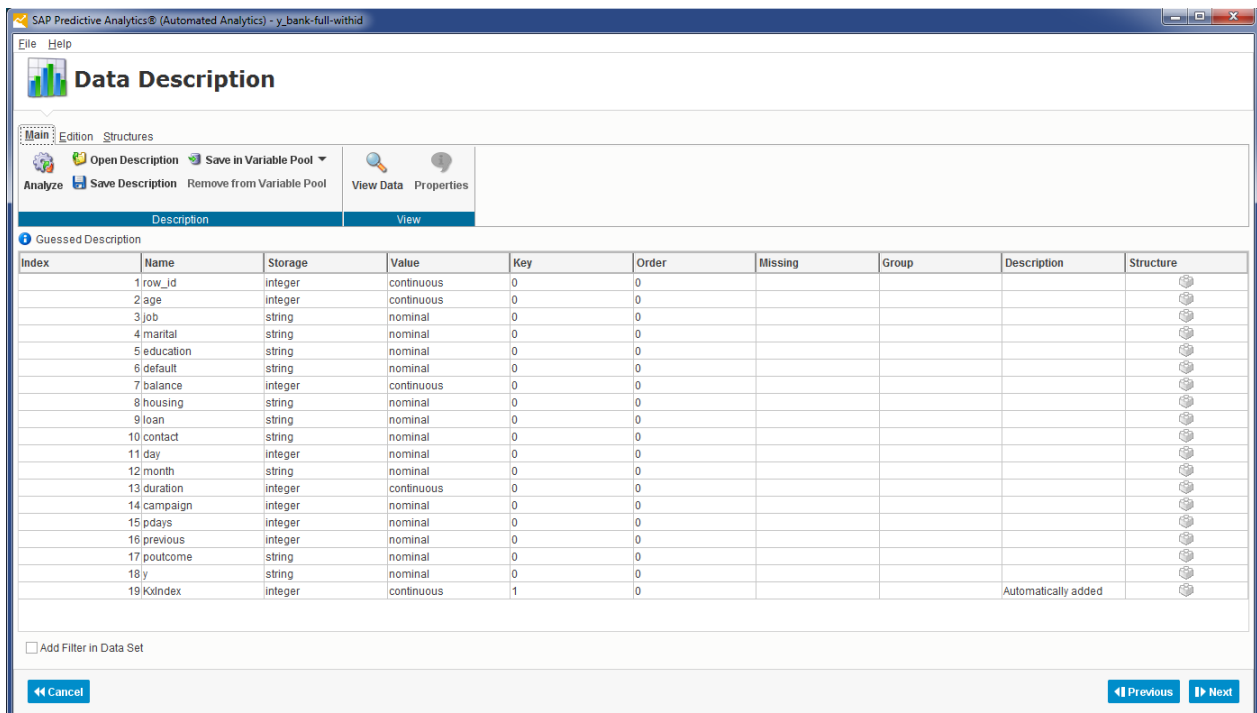
As example, we combine the age and job into one combined column.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Name it “AgeJob”. Set the Type to “Cross” and select the “age” and “job” columns.



Click “OK” and “Close”. This brings you back to the “Data Description” screen.



## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

If you click the “View Data” icon, you will see the new “AgeJob” column.

Automated Analytics Sample Data View

Data Set: bank-full-withid.csv

First Row Index: 1 Last Row Index: 100

Data Statistics Graph

|    | tact | day | month | duration | campaign | pdays | previous  | poutcome | y | KxIndex | AgeJob           |
|----|------|-----|-------|----------|----------|-------|-----------|----------|---|---------|------------------|
| 1  | nawn |     | 5 may | 219      | 1        | -1    | 0 unknown | no       |   | 19      | <60 \"retired... |
| 2  | nawn |     | 5 may | 528      | 1        | -1    | 0 unknown | no       |   | 67      | <60 \"retired... |
| 3  | nawn |     | 5 may | 101      | 1        | -1    | 0 unknown | no       |   | 98      | <60 \"retired... |
| 4  | nawn |     | 5 may | 22       | 1        | -1    | 0 unknown | no       |   | 43      | <60 \"blue-c...  |
| 5  | nawn |     | 5 may | 221      | 1        | -1    | 0 unknown | no       |   | 94      | <60 \"blue-c...  |
| 6  | nawn |     | 5 may | 208      | 1        | -1    | 0 unknown | no       |   | 33      | <60 \"admin...   |
| 7  | nawn |     | 5 may | 583      | 1        | -1    | 0 unknown | no       |   | 93      | <60 \"admin...   |
| 8  | nawn |     | 5 may | 273      | 1        | -1    | 0 unknown | no       |   | 68      | <59 \"mana...    |
| 9  | nawn |     | 5 may | 226      | 1        | -1    | 0 unknown | no       |   | 34      | <59 \"blue-c...  |
| 10 | nawn |     | 5 may | 1042     | 1        | -1    | 0 unknown | yes      |   | 84      | <59 \"admin...   |
| 11 | nawn |     | 5 may | 71       | 1        | -1    | 0 unknown | no       |   | 14      | <58 \"techni...  |
| 12 | nawn |     | 5 may | 355      | 1        | -1    | 0 unknown | no       |   | 47      | <58 \"self-e...  |
| 13 | nawn |     | 5 may | 50       | 1        | -1    | 0 unknown | no       |   | 9       | <58 \"retired... |
| 14 | nawn |     | 5 may | 616      | 1        | -1    | 0 unknown | no       |   | 45      | <58 \"retired... |
| 15 | nawn |     | 5 may | 261      | 1        | -1    | 0 unknown | no       |   | 1       | <58 \"mana...    |
| 16 | nawn |     | 5 may | 225      | 1        | -1    | 0 unknown | no       |   | 31      | <57 \"techni...  |
| 17 | nawn |     | 5 may | 242      | 1        | -1    | 0 unknown | no       |   | 36      | <57 \"techni...  |
| 18 | nawn |     | 5 may | 174      | 1        | -1    | 0 unknown | no       |   | 15      | <57 \"servic...  |
| 19 | nawn |     | 5 may | 180      | 2        | -1    | 0 unknown | no       |   | 90      | <57 \"retired... |
| 20 | nawn |     | 5 may | 173      | 1        | -1    | 0 unknown | no       |   | 95      | <57 \"entrep...  |
| 21 | nawn |     | 5 may | 38       | 1        | -1    | 0 unknown | no       |   | 18      | <57 \"blue-c...  |

Help Close

Note that we were able to combine a numerical column with a nominal column. This is possible as SAP Predictive Analytics transforms the numerical data into multiple subgroups. This subgroup got concatenated with the person's job.

## Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

### Output customer ID with probability

By default, when exporting the probabilities the output does not contain the ID column that identifies the customer. However, often you will need to be able to identify the client in the output.

Create the model as before. When you get to applying the model, click “Advanced Apply Settings”.

**Applying the Model**

Application Data Set

Data Type: Text Files

Folder: C:\Users\I056450\Desktop

Data: bank-full-withid.csv

Generation Options

Generate: Probability & Error Bars

Mode: Apply

Results Generated by the Model

Data Type: Text Files

Folder: C:\Users\I056450\Desktop

Data: term-deposit-probability.csv

Buttons: Cancel, Previous, Apply

In the “General Output” section, tick “Copy Variables”. Select the “row\_id” and hit “OK”.

**Model Advanced Apply Settings**

Advanced Apply Settings

General Outputs

Copy Weight Variable

Copy Variables

Available

Selected

row\_id

User Defined Constant Outputs

| Visibility               | Name          | Storage  | Value               | Key |
|--------------------------|---------------|----------|---------------------|-----|
| <input type="checkbox"/> | Model Name    | string   | y_bank-full-withid  | 0   |
| <input type="checkbox"/> | Build Date    | datetime | 2015-08-08 17:15:17 | 0   |
| <input type="checkbox"/> | Apply Date    | datetime | 2015-08-08 17:16:51 | 0   |
| <input type="checkbox"/> | Model Version | integer  | 1                   | 0   |

Buttons: Cancel, Previous, OK

Hands-On Tutorial SAP Predictive Analytics, Automated Mode: Classification

Then continue as before, and the output will include the ID column.

SAP Predictive Analytics® (Automated Analytics) - y\_bank-full-withid

FileHelp

Applying the Model

Stop

View Type

Copy

Print

Save

Output Name: term-deposit-probability.csv

First Row Index: 1Last Row Index: 100

Data

Statistics

Graph

|    | KxIndex | row_id | y  | rr_y         | proba_rr_y   | bar_rr_y     |
|----|---------|--------|----|--------------|--------------|--------------|
| 1  |         | 1      | no | 0.1534598... | 0.0498061... | 0.8262479... |
| 2  |         | 2      | no | 0.1011557... | 0.0339730... | 0.6195560... |
| 3  |         | 3      | no | 0.0451286... | 0.0258224... | 0.5052350... |
| 4  |         | 4      | no | 0.1265934... | 0.0420584... | 0.7596848... |
| 5  |         | 5      | no | 0.1563053... | 0.0501587... | 0.8386631... |
| 6  |         | 6      | no | 0.1241770... | 0.0412771... | 0.7745867... |
| 7  |         | 7      | no | 0.2771562... | 0.0718238... | 1.1220215... |
| 8  |         | 8      | no | 0.1007592... | 0.0338773... | 0.6143172... |
| 9  |         | 9      | no | 0.0617863... | 0.0274774... | 0.6791833... |
| 10 |         | 10     | no | 0.1439663... | 0.0447874... | 0.7765946... |
| 11 |         | 11     | no | 0.1273262... | 0.0421735... | 0.7556279... |
| 12 |         | 12     | no | 0.2206716... | 0.0544059... | 0.9639066... |
| 13 |         | 13     | no | 0.0434382... | 0.0257204... | 0.5432411... |
| 14 |         | 14     | no | 0.0449538... | 0.0258119... | 0.5097958... |
| 15 |         | 15     | no | 0.0636473... | 0.0276890... | 0.6894703... |
| 16 |         | 16     | no | 0.0838688... | 0.0299885... | 0.5736787... |
| 17 |         | 17     | no | 0.1197830... | 0.0396804... | 0.7678410... |
| 18 |         | 18     | no | 0.0445108... | 0.0257852... | 0.5208654... |
| 19 |         | 19     | no | 0.2293324... | 0.0545687... | 1.0089295... |

Cancel

PreviousNext

© 2015 SAP SE. All rights reserved.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, SAP HANA, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company.

Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase Inc. Sybase is an SAP company.

Crossgate, m@gic EDDY, B2B 360°, and B2B 360° Services are registered trademarks of Crossgate AG in Germany and other countries. Crossgate is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

